EXHIBIT 2

indicate that TFIIB will lie on the opposite face to TFIIA in the preinitiation complex.

TFIIB exists in solution as a monomer[22] and is presumed to bind the TBP–DNA complex as a monomer, but the exact stoichiometry of TFIIB in the TFIIB–TBP complex is not yet known. However, a monomer of the size of the C-terminal domain of TFIIB could interact with both ends of the bent DNA. For comparison, the conserved domain of TBP shown in Fig. 3 is 180 amino acids and the C-terminal domain of TFIIB is 208 amino acids long.

The sugars protected by TFIIB both upstream and downstream of the TATA box border the minor groove in the TBP–DNA model. This suggests that TFIIB, like TBP, interacts at least in part with DNA in the minor groove. Consistent with this prediction, TFIIB binding did not protect any G residues from methylation by dimethyl sulphate at the G+C-rich Ad-MLP (data not shown). TFIIB does not appear to make essential base-specific contracts with the major groove of the TATA box, as both TBP[12] and TBP–TFIIB form stable complexes with a (dI-dC)-substituted MLP TATA box (data not shown).

Our model for the binding of TFIIB to TBP–DNA revises previous models. Most describe TFIIB interactions with DNA only downstream of the TATA box[1,3,6]. Our results show that, in addition to the C-terminal stirrup of TBP, a critical element of the TFIIB target is DNA bent in a specific orientation. It is unlikely that a monomer the size of TFIIB could interact with DNA on both sides of the TATA box if the DNA is not bent closer on itself by TBP. TBP may need to bend DNA when it binds in order to support the assembly of preinitiation complexes. TBP bound to distorted DNA is likely to be an important target for other components of the preinitiation complex[6,15]. □

Received 10 May; accepted 10 July 1995.

1. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. *Cell* **56**, 549–561 (1989).
2. Maldonado, E., He, I., Cortes, P., Weis, L. & Reinberg, D. *Molec. cell. Biol.* **10**, 6335–6347 (1990).
3. Choy, B. & Green, M. R. *Nature* **366**, 531–536 (1993).
4. Roberts, S. G. E. & Green, M. R. *Nature* **371**, 717–720 (1994).
5. Malik, S., Lee, D. K. & Roeder, R. G. *Molec. cell. Biol.* **13**, 6253–6259 (1993).
6. Coulombe, B. L. & Greenblatt, J. J. *biol. Chem.* **269**, 19962–19967 (1994).
7. Barberis, A., Muller, C. W., Harrison, S. C. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5628–5632 (1993).
8. Buratowski, S. & Zhou, H. *Proc. natn. Acad. Sci. U.S.A.* **90**, 5633–5637 (1993).
9. Ha, I. *et al. Genes Dev.* **7**, 1021–1032 (1993).
10. Yamashita, S. *et al. Science* **261**, 463–466 (1993).
11. Tufilus, T. D. *Nature* **333**, 913–664 (1988).
12. Starr, D. B. & Hawley, D. K. *Cell* **67**, 1231–1240 (1991).
13. Lee, D. K., Horikoshi, M. & Roeder, R. G. *Cell* **67**, 1241–1250 (1991).
14. Kimball, A. S., Lee, J., Jayaram, M. & Tullius, T. D. *Biochemistry* **32**, 4698–4701 (1993).
15. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. *Nature* **365**, 512–520 (1993).
16. Kim, J. L., Nikolov, D. B. & Burley, S. K. *Nature* **365**, 520–527 (1993).
17. Nikolov, D. B. *et al. Nature* **360**, 40–46 (1992).
18. Kim, J. L. B. & Burley, S. K. *Nature struct. Biol.* **1**, 638–653 (1994).
19. Kim, T. K. *et al. Nature* **365**, 252–255 (1994).
20. Buratowski, S. & Zhou, H. *Science* **255**, 1130–1132 (1992).
21. Lee, D. K., Dejong, J., Hashimoto, S., Horikoshi, M. & Roeder, R. G. *Molec. cell. Biol.* **12**, 5189–5196 (1992).
22. He, I., Lane, W. S. & Reinberg, D. *Nature* **353**, 669–665 (1991).
23. Clark, J. M. *Nucleic Acids Res.* **18**, 9677–9686 (1988).
24. Roberts, S., Colbert, T. & Hahn, S. *Genes Dev.* **9**, 832–833 (1995).

ACKNOWLEDGEMENTS. We thank D. Reinberg and Y. Kim for human TFIIB; J. Geiger for truncated yeast TBP; R. Ebright and D. Reinberg for communication of unpublished results; D. Auble, Y. Kim, J. Geiger, R. Ebright, P. Hopkins and T. Tullius for advice and discussion; members of the Hahn laboratory, D. Reinberg, R. Ebright, Y. Kim, J. Geiger and P. Sigler for comments on the manuscript; and B. Stoddard for help with graphics. This work was supported by a grant from the NIH and a Leukemia Society Scholar award to S.H.

Crystal structure of *Thermus aquaticus* DNA polymerase

Youngsoo Kim*†, Soo Hyun Eom*‡, Jimin Wang*, Dae-Sil Lee§, Se Won Suh‡ & Thomas A. Steitz*‖

* Department of Molecular Biophysics and Biochemistry,
‖ Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06520-8114, USA
‡ Department of Chemistry, Center for Molecular Catalysis, Seoul National University, Seoul 151-742, Korea
§ Korea Research Institute of Bioscience and Biotechnology, Korean Institute of Science and Technology, Taejon 305-333, Korea

THE DNA polymerase from *Thermus aquaticus* (*Taq* polymerase), famous for its use in the polymerase chain reaction, is homologous to *Escherichia coli* DNA polymerase I (pol I) (ref. 1). Like pol I, *Taq* polymerase has a domain at its amino terminus (residues 1–290) that has 5′ nuclease activity and a domain at its carboxy terminus that catalyses the polymerase reaction. Unlike pol I, the intervening domain in *Taq* polymerase has lost the editing 3′–5′ exonuclease activity. Although the structure of the Klenow fragment of pol I has been known for ten years[2], that of the intact pol I has proved more elusive. The structure of *Taq* polymerase determined here at 2.4 Å resolution shows that the structures of the polymerase domains of the thermostable enzyme and of the Klenow fragment are nearly identical, whereas the catalytically critical carboxylate residues that bind two metal ions are missing from the remnants of the 3′–5′-exonuclease active site of *Taq* polymerase. The first view of the 5′ nuclease domain, responsible for excising the Okazaki RNA in lagging-strand DNA replication, shows a cluster of conserved divalent metal-ion-binding carboxylates at the bottom of a cleft. The location of this 5′-nuclease active site some 70 Å from the polymerase active site in this crystal form highlights the unanswered question of how this domain works in concert with the polymerase domain to produce a duplex DNA product that contains only a nick.

Crystals of intact *Taq* polymerase (Table 1) diffract to 2.4 Å resolution at −165 °C using synchrotron radiation. The structure was solved initially from a 3.3 Å-resolution electron-density map, phased by multiple heavy-atom isomorphous replacement and improved by solvent flattening using a manually drawn envelope (Fig. 1 and Table 1). Although the polymerase domain shows a 51% amino-acid sequence identity with that of pol I (ref. 1), knowledge of the Klenow fragment (KF) structure did not help in the early stages of phasing, because even this conserved portion contributed too small a fraction to the X-ray scattering. The coordinates of *Taq* polymerase have been partially refined to an *R*-factor of 22.9% (R_{free} = 32.2%), with r.m.s. bond and angle deviations of 0.011 Å and 1.79°, respectively, for all data between 10 and 2.4 Å resolution. The tip of the 'thumb' in the polymerase domain is disordered and there are several regions in the 5′ nuclease domain where the electron density is discontinuous, presumably because of disordering of loops: these regions include residues 12–13, 68–89, 151–172 and 199–202. Furthermore, the residues from 172 to 233 are built here as polyalanine, again for reasons of poorly ordered electron density. The strung-out arrangement of the three domains in *Taq* polymerase results in an unusually elongated molecule that is 130 Å long in this crystal form (Fig. 2*a*).

Comparison of the structure of KF with the corresponding parts of the *Taq* polymerase structure shows, as expected from the sequence comparisons, that the polymerase domains are very nearly identical, whereas the 3′–5′ exonuclease domains differ extensively (Fig. 2). Least-squares superposition of 353 of 407 corresponding α-carbon atoms in the polymerase domains resulted in an r.m.s. difference of 1.2 Å. By contrast, only 101 of 194 α-carbon atoms in the KF 3′–5′ exonuclease domain could

† Present address: Department of Industrial Chemistry, Yeungnam University, Taegu 712-749, Korea.

TABLE 1 Experimental X-ray data and heavy-atom refinement

	Resolution (Å)	Completion (%)	No. unique reflections	R^*_{sym} (%)	$R^†_{diff}$ (%)	Phasing‡§ power	Mean figure of merit
Native I	2.6	80.0	28,179	6.7	—	—	0.686
Native II	2.4	90.0	40,493	4.6	—	—	
10 mM TMLA‖	2.5	80.0	31,443	6.9	13.4	1.0	
1 mM CH_3HgCl	4.0	99.4	9,814	6.9	8.7	0.85	
Sat. baker's dimercurial	3.0	95.5	22,127	11.4	11.3	0.54	
1 mM K_2PtCl_4	3.3	97.1	17,447	11.8	27.3	0.53	
1 mM K_2PtI_4	4.0	70.2	6,987	17.6	39.7	0.66	
1 mM Ch_3HgCl/10 mM TMLA‖	3.3	92.0	16,541	11.8	24.9	1.10	

Additional data sets were collected that extended to between 4 and 5 Å resolution for all heavy-atom derivatives (data not shown), and their phasing information was combined with the data shown in the table. These supplementary data produced a map with a better-defined solvent boundary and an improved connectivity of the main-chain backbone. Crystals of *Taq* polymerase were grown at 22 °C in hanging drops containing 3 µl protein solution (10 mg ml^{-1} *Taq* polymerase in 18 mM Tris-HCl, pH 8.2, 0.09 mM EDTA, 0.9 mM DTT, 90 mM KCl, 9% (v/v) glycerol and 0.7 % (w/v) β-octyl glucoside) and 3 µl reservoir solution (15% (w/v) PEG8000, 60 mM ammonium sulphate, 2 mM DTT, 0.2% (w/v) sodium azide and 100 mM sodium citrate, pH 5.5)[21]. The crystals belong to space group $P3_121$ and have unit cell dimensions of $a=b=108.0$ Å, $c=171.2$ Å, $\alpha=\beta=90°$ and $\gamma=120°$. The presence of one molecule per asymmetric unit gives a crystal volume per protein mass (V_m) of 3.54 $Å^3$ per dalton and a solvent content of 65% by volume[22]. In order to make heavy-atom derivatives with mercurials, wild-type *Taq* polymerase (no cysteine residues) was mutated by site-directed mutagenesis to introduce three consecutive cysteines at positions 575 to 577. Crystals of this protein were used for all native and derivative data sets. Crystals were flash-frozen at −165 °C after first transferring to a stabilizing solution containing 40 mM sodium citrate, pH 5.5, 10% (v/v) glycerol, 100 mM KCl, 0.4% (w/v) β-octyl glucoside and 31% (w/v) PEG8000 for 24 h, and then to a second stabilizing solution containing 20 mM HEPES, pH 7.4, 10% (v/v) glycerol, 100 mM KCl, 0.4% (w/v) β-octyl glucoside and 33% (w/v) PEG8000 for 36 h. A 2.6-Å native data set collected at BL-6A2 of the Photon Factory was used in the initial survey for heavy-atom derivatives, but not subsequently. Native I and four derivative data sets were collected at the CHESS A1 beam line ($\lambda=0.908$ Å) equipped with a CCD camera detector. The native II data set was collected at the X12C beam line of the National Synchrotron Light Source (NSLS) at the Brookhaven National Laboratory ($\lambda=1.00$ Å equipped with the MAR X-ray detector system). Two derivative data sets were collected on an R-AXIS IIC X-ray detector system mounted on a Rigaku-200 rotating anode. All data were reduced using DENZO and scaled using SCALEPACK (programs written by Z. Otwinowski). The position of the *Taq* polymerase in the crystal was solved by molecular replacement at 4 Å resolution using a model of the polymerase domain that contained 45% of the scattering mass of *Taq* polymerases and was based on the structure of Klenow fragment refined at 2.5 Å resolution (J. Jaeger, D. Carrill and T.A.S., unpublished result). The rotation function search was done using MERLOT[23]; the Patterson correlation refinement and the translation function in X-PLOR[24]. Phases calculated from the model allowed location of bound heavy atoms by difference-Fourier syntheses at 4 Å resolution but were not used directly in the structure determination. An MIR electron density map calculated at 3.3 Å resolution using these derivatives refined with ML-PHARE[25] was improved by solvent flattening, histogram matching and phase combination using SQUASH[26]. A polyalanine model fitted to this map allowed computation of a molecular envelope around the model with a 5 Å radius for each atom using the program O (ref. 27). Further solvent flattening with this manually drawn envelope using SQUASH and interactive heavy-atom refinement against flattened phases improved the map quality. Refinement of the structure built into this map, including simulated annealing, position refinement and manual model rebuilding was done against data from 10 to 2.4 Å resolution. The present structure includes 776 amino acids, one β-octyl glucoside, and 297 water molecules. Fifty-six amino acids are not visible and sixty are modelled by polyalanine.

* $R_{sym}=\Sigma|I-\langle I\rangle|/\Sigma I$, where I is the observed intensity and $\langle I\rangle$ is the average intensity from multiple measurements.

† $R_{diff}=\Sigma||F_{PH}|-|F_P||/\Sigma|F_P|$

‡ Data beyond 3.3 Å resolution were not included in the phase refinement of heavy-atom statistics for any derivatives.

§ Phasing power, F_H/ε=r.m.s. (F_H)/r.m.s. (lack of closure), where F_H is the calculated heavy-atom structure factor.

‖ TMLA, trimethyl lead acetate. A crystal was soaked in 1 mM CH_3HgCl for 20 h and then transferred to 10 mM TMLA solution to be soaked for 72 h.

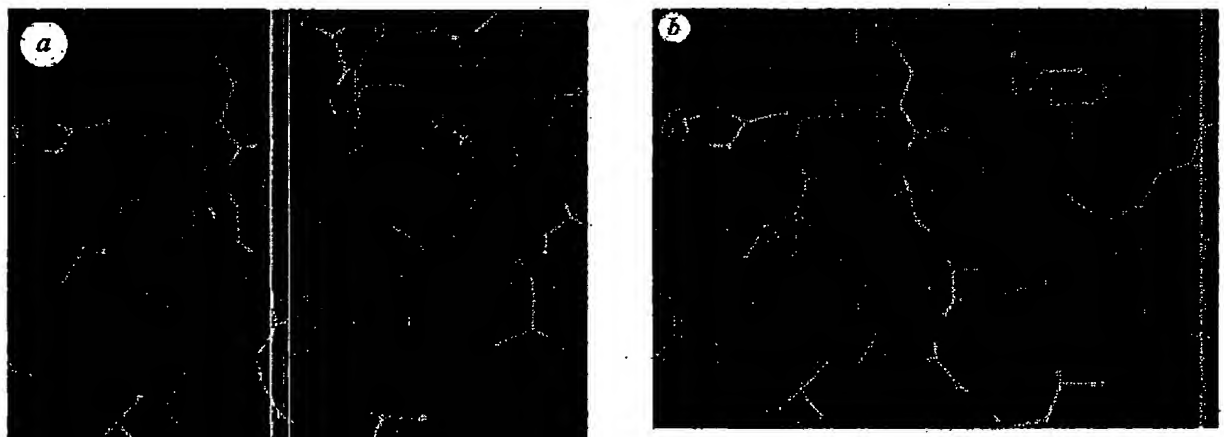

FIG. 1 Electron density maps with the refined model superimposed. *a*, Solvent-flattened multiple isomorphous replacement (MIR) map at 3.3 Å resolution; *b*, 2.4 Å-resolution deletion map calculated using $2F_o-F_c$ as amplitudes and using phases calculated from the structure omitting the portion of the model shown.

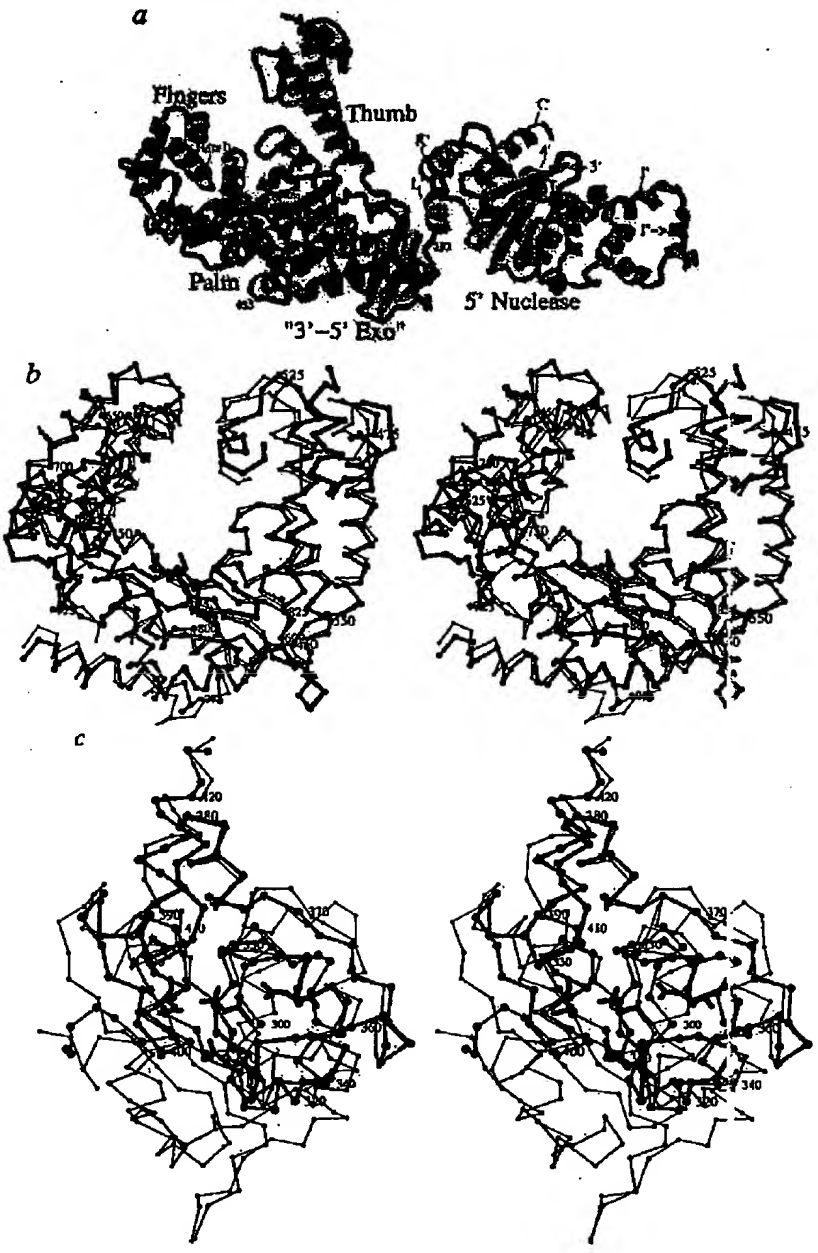

FIG. 2 *Taq* polymerase structure and comparison with that of KF. *a*, Helix-and-arrow overall schematic ribbon representation of *Taq* polymerase drawn using rendered MOLSCRIPT[28,29]. α-Helices are represented as helical ribbons and β-strands as arrows. Helices and strands are lettered and numbered as in KF for the 3′–5′ exonuclease and polymerase domains and lettered and numbered with primes in the 5′ nuclease (previously called 5′–3′ exonuclease) domain. The 5′ nuclease domain at the N terminus is orange and yellow; its active site is marked by a red Zn^{2+} and two blue Mn^{2+} ions. The portion of this domain whose side chains have not been positioned is yellow (residues 172–234 and 1–12). The vestigial 3′–5′ exonuclease domain is red and the polymerase domain is divided into green thumb, blue finger and purple palm subdomains. The active site Asp 610, Asp 785 and Glu 786 are in dark green. *b*, Superposition of KF and *Taq* polymerase. Stereo α: Cα backbone of the KF polymerase domain (thin bonds) superimposed on the corresponding atoms (thick bonds) of the *Taq* polymerase, which are numbered. The three catalytic carboxylate side chains are shown in ball and stick representation at the bottom of the cleft. *c*, Superposition of 131 α-carbon atoms of the 3′–5′ exonuclease domain of KF (thin bonds) on the corresponding atoms of *Taq* polymerase. The four catalytic carboxylates in KF are shown. *d*, Structure based alignment of the sequences of the 3′–5′ exonuclease domain of KF on the corresponding domain of *Taq* polymerase. The unaligned residues are shown as dots in *E. coli* polymerase I (E.c.) and the unpaired missing residues are shown as blank in *Taq* polymerase (T.aq). The amino-acid sequence numbers of *Taq* DNA polymerase secondary structure elements are as follows: 5′ nuclease domain: 1′(3–7) 2′(12–17) A′(18–29) B′(42–57) 3′(60–67) C′(91–106) 4′(108–113) D′(119–132) 5′(134–139) E′(143–148) 6′(175–178) F′(179–183) G′(189–198) H′(203–213) h′(217–221) I′(225–232) J′(235–246) K′(266–276) L′(281–289); 3′–5′ exonuclease domain: 1(294–298) A(—) 2(305–312) 3(322–328) 4(330–336) B(338–344) 5(347–351) C(353–362) 5a(368–372) D(373–380) E(387–394) F(402–422); polymerase domain: 3(424–447) 6(448–452) H(453–477) Ha(487–496) Hb(515–521) I(527–552) 7(559–598) 8(572–576) J(581–584) K(589–598) 9(603–613) L(614–623) M(626–634) N(638–648) O(656–670) Oa(675–723) Ob(688–699) P(702–717) 10(718–723) 11(724–729) Q(740–774) 12(776–784) 13(785–792) R(795–810) 14(816–825) Re(826–831).

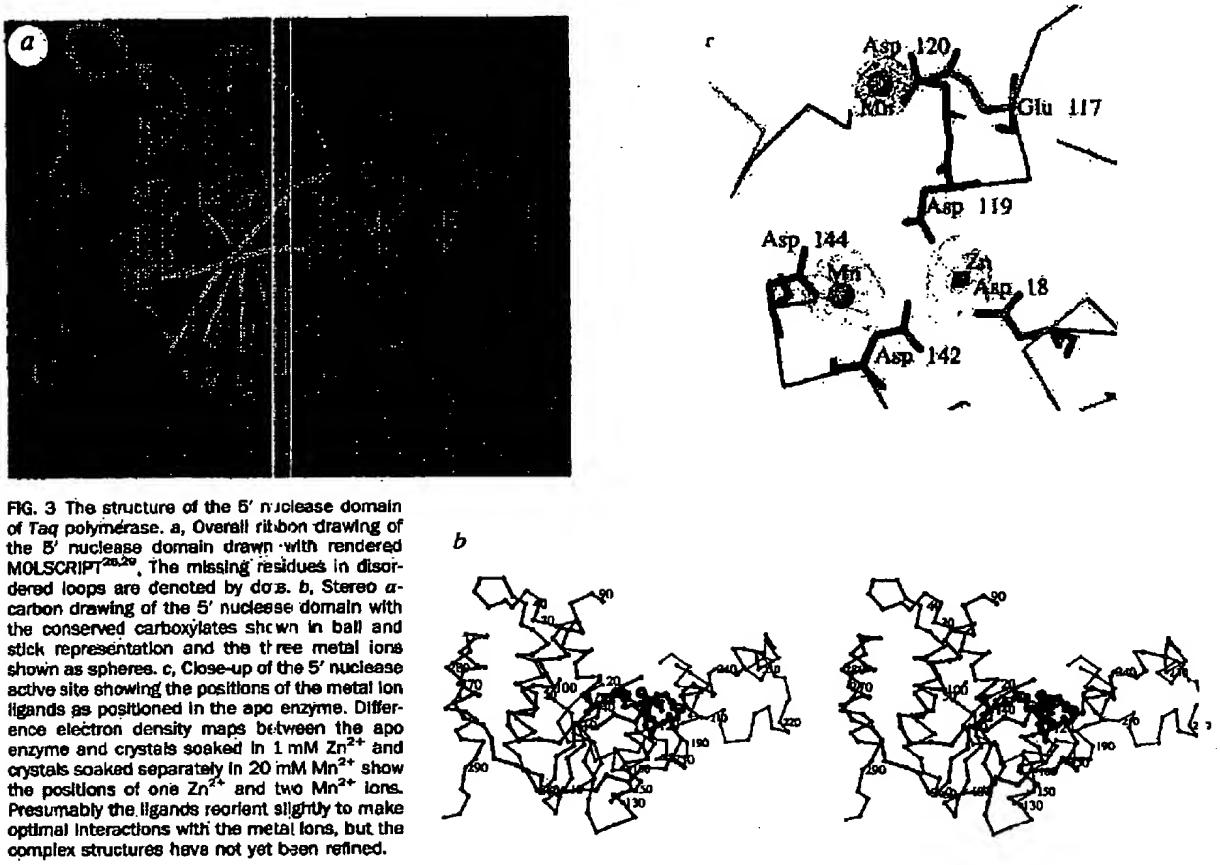

FIG. 3 The structure of the 5′ nuclease domain of *Taq* polymerase. *a*, Overall ribbon drawing of the 5′ nuclease domain drawn with rendered MOLSCRIPT[28,29]. The missing residues in disordered loops are denoted by dots. *b*, Stereo α-carbon drawing of the 5′ nuclease domain with the conserved carboxylates shown in ball and stick representation and the three metal ions shown as spheres. *c*, Close-up of the 5′ nuclease active site showing the positions of the metal ion ligands as positioned in the apo enzyme. Difference electron density maps between the apo enzyme and crystals soaked in 1 mM Zn^{2+} and crystals soaked separately in 20 mM Mn^{2+} show the positions of one Zn^{2+} and two Mn^{2+} ions. Presumably the ligands reorient slightly to make optimal interactions with the metal ions, but the complex structures have not yet been refined.

be superimposed on the corresponding atoms in the 131-residue domain in the *Taq* enzyme to give an r.m.s. difference of 1.6 Å. One major difference in the overall structure of the 3′–5′ domain of *Taq* polymerase as compared with that of KF is the deletion of four loops of lengths between 8 to 27 residues. In KF, these loops pack together on one side of the 3′–5′ exonuclease domain (Fig. 2*b*). Furthermore, all four of the carboxylates (D424, D501, D355, E357) known to be essential for divalent metal binding and catalysis in the 3′–5′ exonuclease domain of KF[3,4] have been replaced by residues incapable of binding metal ions (L356, R405, G308, V310) in the vestigial 3′–5′ exonuclease domain of *Taq* polymerase. Although the 3′–5′ exonuclease catalytic site has been destroyed and the size of the domain reduced, the contact with the pol domain and the distance between the polymerase and 3′–5′ exonuclease domains remains similar in the two homologous polymerases.

The 5′-nuclease domain forms a structure that is separate from the other two domains (Fig. 2), with only 850 Å^2 of surface area in contact with the 3′–5′ exonuclease domain, consistent with this domain's ability to function after its proteolytic removal from the rest of the protein (J. B. Dahlberg, personal communication, and ref. 5). It has a deep cleft that contains at its bottom the conserved carboxylates shown here to ligate divalent metal ions. A central β-sheet lies at the heart of the domain and is flanked on both sides by assemblies of five and six α-helices which form the walls of the active-site cleft (Fig. 3).

Alignment of the amino-acid sequences of six 5′ nuclease domains from DNA polymerases in the pol I family show six highly conserved sequence motifs containing ten conserved acidic residues[6]. Seven of these residues (Asp 18, Asp 67, Glu 117, Asp 119, Asp 120, Asp 142 and Asp 144) cluster within a sphere of 7 Å radius, two (Asp 188 and Asp 191) lie in a region built as polyalanine, and one (Glu 76) occurs in a completely disordered loop. Crystallographic data from crystals soaked in divalent metal ions show that some of these carboxylates serve to ligate as many as three divalent metal ions (Fig. 3*c*). A difference Fourier using data from crystals soaked in 20 mM Mn^{2+} shows two peaks, corresponding to a metal-ion site III that is interacting with Glu 117, Asp 120, and possibly Asp 119, and a metal-ion site II that is interacting with Asp 142 and Asp 144. Soaking crystals in 1 mM Zn^{2+} reveals a divalent metal-ion-binding site I, whose ligands appear to be Asp 18, Asp 119, Asp 142 and perhaps His 21. Electron density maps of the apo protein show some density at metal-ion site I, which may indicate binding of a partially substituted Zn^{2+} ion. Sites I and II are separated by about 5 Å, whereas these two sites are each about 10 Å from site III.

As we do not yet have the structures of either substrate or product complexes with the enzyme, a firm mechanism for 5′-nuclease reaction[7] cannot be proposed. However, a mechanism of phosphoryl transfer is becoming apparent in an increasing number of enzymes in which two divalent metal ions are involved. This two-metal-ion mechanism was suggested initially for the 3′–5′ exonuclease of KF[8] and is supported by structural, mutagenic and kinetic studies[3,4,9]. There is evidence that the enzymes alkaline phosphatase[10], pyrophosphatase[11,12], RNase H (refs 13, 14) and polymerase, to name a few, use a similar mechanism, as may[15–17] other enzymes containing conserved, catalytically essential carboxylates (such as ruvC[18] and HIV integrase[19])[20]. In this mechanism, the two metal ions are

generally 4 Å apart, interact directly with the scissile phosphate, stabilize the pentacovalent intermediate, and generate the attacking hydroxide ion, as well as facilitating the departure of the 3′ oxyanion[8]. If such a two-metal-ion mechanism is relevant to the 5′ nuclease, the possible role, if any, of the more distant site III metal ion can only be guessed.

The source of the thermal stability of *Taq* polymerase is not obvious from structural comparison with KF, but the number of hydrogen bonds has increased by four, and two salt bridges between subdomains in the polymerase domain become hydrophobic; the ratio of leucine to isoleucine has increased by 4.4-fold, and arginine to lysine by 1.3-fold, which may result from the higher G+C content of the leucine and arginine codons (giving a more thermostable DNA), rather than an effect on the protein.

An important question concerning the pol I family of enzymes is how the polymerase and 5′-nuclease active sites work together to generate a duplex DNA product containing only a nick; the present structure raises at least as many questions as it answers, because we observe that these two active sites are separated by over 70 Å. The unusually elongated shape of the molecule seen here led us to examine its overall fold in solution. Preliminary measurements of the radius of gyration (R_g) of *Taq* polymerase by solution X-ray-scattering methods yield an experimental value of R_g that is substantially smaller than that calculated from the coordinates of the crystal structure (S.H.E. *et al.*, unpublished observations). Thus the 5′ nuclease domain is not positioned in solution as shown in Fig. 2, but must be located much closer to the centre of mass of the Stoffel fragment. Presumably its orientation in these crystals is adventitious and governed by crystal-packing interactions. Two packing interactions between the 5′ nuclease and neighbouring molecules bury 1,100 and 1,466 Å^2 of solvent-accessible area, larger than the intramolecular interaction surface. A structural basis for understanding how these two activities work together must await the crystal structure of a complex with the appropriately nicked DNA substrate. □

Received 11 May; accepted 3 July 1995.

1. Lawyer, F. C. *et al. J. biol. Chem.* **264**, 6427–6427 (1989).
2. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. *Nature* **313**, 762–766 (1985).
3. Derbyshire, V. *et al. Science* **240**, 199–201 (1988).
4. Derbyshire, V., Grindley, N. D. F. & Joyce, C. M. *EMBO J.* **10**, 17–24 (1991).
5. Lundquist, R. C. & Olivera, B. M. *Cell* **31**, 53–60 (1982).
6. Gutman, P. D. & Minton, K. W. *Nucleic Acids Res.* **21**, 4406–4407 (1993).
7. Lyamichev, V., Brow, M. A. D. & Dahlberg, J. E. *Science* **260**, 778–783 (1993).
8. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. & Steitz, T. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8924–8928 (1988).
9. Beese, L. S. & Steitz, T. A. *EMBO J.* **10**, 25–33 (1991).
10. Kim, E. & Wyckoff, H. W. *J. molec. Biol.* **218**, 449–464 (1991).
11. Kankare, J. *et al. Protein Engng* **7**, 823–830 (1994).
12. Teplyakov, A. *et al. Protein Sci.* **7**, 1098–1107 (1994).
13. Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. *Science* **249**, 1398–1405 (1990).
14. Davies, J. F., Hostomska, Z., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. *Science* **252**, 88–95 (1991).
15. Steitz, T. A. *Curr. Opin. struct. Biol.* **3**, 31–38 (1993).
16. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. *Science* **264**, 1891–1903 (1994).
17. Steitz, T. A., Smerdon, S. J. Jäger, J. & Joyce, C. M. *Science* **266**, 2022–2025 (1994).
18. Ariyoshi, M. *et al. Cell* **78**, 1063–1072 (1994).
19. Dyda, F. *et al. Science* **266**, 1981–1986 (1994).
20. Yang, W. & Steitz, T. A. *Structure* **3**, 131–134 (1995).
21. Eom, S. H. *et al. Acta crystallogr.* **D51** (in the press).
22. Matthews, B. C. W. *J. molec. Biol.* **33**, 491–497 (1968).
23. Fitzgerald, P. M. D. *J. appl. Crystallogr.* **21**, 273–278 (1988).
24. Brünger, A. T. *X-PLOR, version 3.1* (Yale Univ. Press, New Haven, CT, 1992).
25. Otwinowski, Z. *Ml—PHARE CCP4 Proc.* 80–86 (Daresbury Laboratory, Warrington, 1991).
26. Zhang, K. Y. J. & Main, P. *Acta crystallogr.* **A46**, 377–381 (1990).
27. Jones, T. A. & Kjeldgaard, M. *O version 5.9* (Department of Molecular Biology, BMC, Uppsala Univ., Sweden, 1993).
28. Kraulis, P. J. *J. appl. Crystallogr.* **24**, 946–950 (1991).
29. Merritt, E. A. & Murphy, M. E. P. *Acta crystallogr.* **D50**, 869–873 (1994).

ACKNOWLEDGEMENTS. We thank Y. Sakabe and A. Nakagawa of the BL-6A2 beamline at the Photon Factory, Japan; S. Ealick of the A1 beam line at the CHESS; R. Sweet of the X12C beam line at the NSLS of the Brookhaven National Laboratory for access and help; A. M. Friedman for help with data collection at the A1 beam line; and J. M. Friedman, D. Jeruzalmi and P. Pepin for additional help. This work was supported by a grant from the American Cancer Society (to T.A.S.) by the Center for Molecular Catalysis (S.W.S.), and by the Korea Ministry of Science and Technology (D.S.L.).

ERRATUM

Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA

Yousif Shamoo, Alan M. Friedman, Mark R. Parsons, William H. Konigsberg & Thomas A. Steitz

Nature 376, 362–366 (1995)

An error in the production process resulted in Fig. 1*a* and *b* of the paper by Kim *et al.* on page 613 of this issue being substituted for Fig. 2*a* and *b* of the earlier paper by Shamoo *et al.* The correct panels of Fig. 2 are shown here.

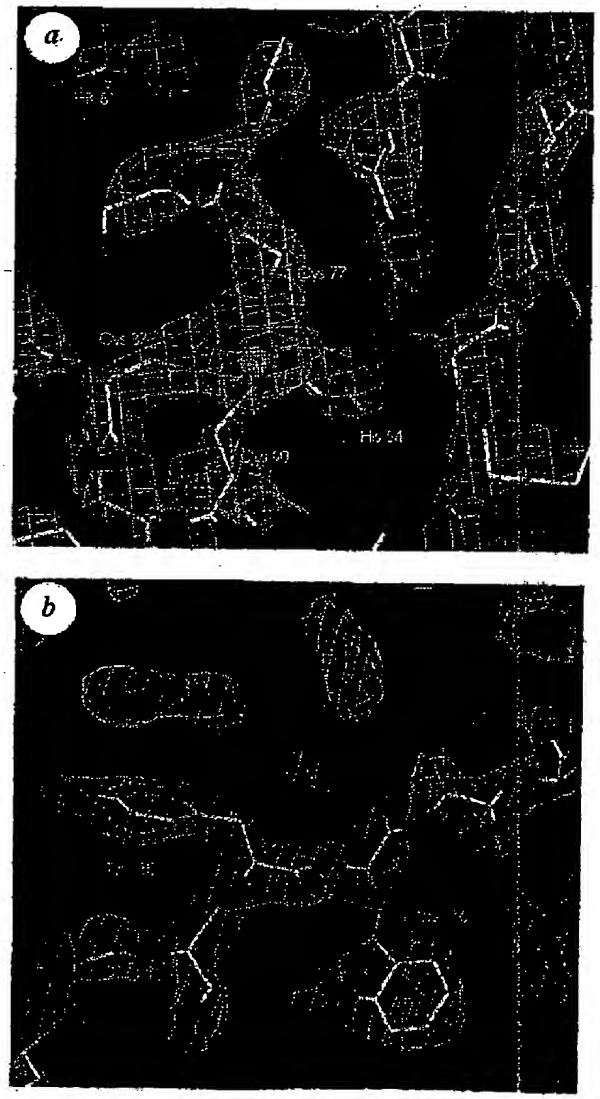

FIG. 2 *a*, Experimental electron density map to 3.1 Å contoured at 1.5 σ and calculated using the combined MAD and MIR phases that have been solvent flattened. The coordination of the Zn^{2+} ion (yellow) is tetrahedral with His 64, Cys 77, Cys 87 and Cys 90 as ligands. *b*, $2F_o-F_c$ electron density map contoured at 1.3 σ showing a stretch of β-strand 4 that includes the current partly refined model and all the data to 2.2 Å.

EXHIBIT 3

Mol. Cells, Vol. 7, No. 6, pp. 769-776

Cloning and Analysis of the DNA Polymerase-encoding Gene from *Thermus filiformis*

Seung Eun Jung, Jeong Jin Choi, Hyun Kyu Kim and Suk-Tae Kwon*
Department of Genetic Engineering, Sung Kyun Kwan University, Suwon 440-746, Korea

(Received on August 14, 1997)

The gene encoding *Thermus filiformis* (*Tfi*) DNA polymerase was cloned and its nucleotide sequence was determined. The primary structure of *Tfi* DNA polymerase was deduced from its nucleotide sequence. *Tfi* DNA polymerase is comprised of 833 amino acid residues and its molecular mass was determined to be 93,890 Da. The deduced amino acid sequence of *Tfi* DNA polymerase showed a high sequence homology to *E. coli* DNA polymerase I-like DNA polymerases: 78.5% homology to *Taq* DNA polymerase, 78.4% to *Tca* DNA polymerase, and 41.8% to *E. coli* DNA polymerase I. An extremely high sequence identity was observed in the region containing polymerase activity. The G+C content of the coding region for the *Tfi* DNA polymerase gene was 68.5%, which was higher than that of the chromosomal DNA (65%). The G+C contents in the first, second, and third positions of the codons used were 71.8%, 40.9%, and 92.7% respectively. Codon usage in *Tfi* DNA polymerase was heavily biased towards the use of G+C in the third position. Rare codons with U or A as the third base were sometimes used to avoid using GA(A/T)TC and TCGA sequences, as they are recognition sites for the restriction endonucleases *Tfi*I and *Taq*I.

DNA polymerase is one of the most important enzymes for DNA repair and replication in living cells. Many different DNA polymerase genes have been cloned and sequenced. Their deduced amino acid sequences have been reported from nucleotide sequence data (Joyce *et al.*, 1982; Lawyer *et al.*, 1989; Lopez *et al.*, 1989). The amino acid sequences of these DNA polymerases have been aligned and partial homologous regions have been identified (Bernad *et al.*, 1989; Blanco *et al.*, 1991; Ito and Braithwaite, 1991). On the basis of segmental similarities in the amino acid sequences, DNA polymerases have been classified into two major groups represented by *E. coli* DNA polymerase I-like prokaryotic DNA polymerases and DNA polymerase α-like prokaryotic and eukaryotic DNA polymerases (Bernad *et al.*, 1989; Blanco *et al.*, 1991). A classification of DNA polymerases into families A, B, and C according to the homology of the amino acid sequence with *E. coli* DNA polymerase I, II, and III, respectively, has been proposed (Ito and Braithwaite, 1991).

We are interested in cloning genes coding for thermostable DNA polymerases, which are useful for polymerase chain reaction (PCR). Recently, PCR has become a powerful method for the identification and amplification of genes, their direct sequencing, and clinical diagnosis. The thermostable DNA polymerase is the key ingredient of PCR. Early experiments used the thermolabile Klenow fragment, which had to be added every cycle (Saiki *et al.*, 1985). The introduction of thermostable DNA polymerase allowed the automation of the process (Saiki *et al.*, 1988). Accordingly, thermostable DNA polymerase was much more stable and suitable in thermocycles during PCR (Erlich, 1989; Saiki *et al.*, 1988).

The purification procedures and properties of thermostable DNA polymerases have been reported for thermophilic bacteria in the genus *Thermus* such as *T. aquaticus* YT-1 (Chien *et al.*, 1976), *T. ruber* (Kaledin *et al.*, 1980), *T. flavus* (Kaledin *et al.*, 1982), *T. thermophilus* HB8 (Ruettimann *et al.*, 1985) and *T. caldophilus* GK24 (Park *et al.*, 1993). However, no information is available on properties of thermostable DNA polymerase from *T. filiformis*.

T. filiformis was isolated from a New Zealand hot spring and was described as a member of the genus *Thermus* by Hudson *et al.* (1987). *T. filiformis* always forms long filaments consisting of chains of cells and so can be distingushed from other *Thermus*

* To whom correspondence should be addressed.

The abbreviations used are: PCR, polymerase chain reaction; *Taq* DNA polymerase, DNA polymerase isolated from *Thermus aquaticus* YT-1; *Tca* DNA polymerase, DNA polymerase isolated from *Thermus caldophilus* GK24; *Tfi* DNA polymerase, DNA polymerase isolated from *Thermus filiformis*. The nucleotide sequence data reported in this paper will appear in the GenBank nucleotide sequence databases with the accession number AF030320.

strains in morphology.

In this paper we report (i) the cloning of the gene for *Tfi* DNA polymerase, (ii) the nucleotide sequence of the *Tfi* DNA polymerase gene and its deduced amino acid sequence, (iii) comparison of the amino acid sequence of *Tfi* DNA polymerase with those of other *E. coli* DNA polymerase I-like DNA polymerases, and (iv) the analysis of the gene.

Materials and Methods

Bacterial strains and culture conditions

T. filiformis (ATCC 43280) (Hudson *et al.*, 1987) cells were prepared as described by Ramaley and Hixson (1970). *E. coli* strain MV1184 (Sambrook *et al.*, 1989) was used as the host for plasmid prepartions and was grown in LB medium supplemented with 0.1% glucose. Ampicillin (100 μg/ml) was added when needed. The *E. coli* cells were grown at 37 ℃. Plates were solidified with 1.5% agar.

Enzymes and reagents

Tfi restriction endonuclease was purchased from New England Biolabs, Inc. T4 DNA ligase, polynucleotide kinase, DNA molecular weight marker X, and other restriction enzymes were purchased from Boehringer Mannheim GmbH. *Taq* DNA polymerase was prepared as described previously (Kwon *et al.*, 1991). An oligo labeling kit and radioactive nucleotides were purchased from Amersham, and Deaza^{35} Sequencing™ Mixes, plasmids pUC18/19 (Norrander *et al.*, 1983), and pBluescript®II SK+/− (Alting-Mees and Short, 1989; Short *et al.*, 1988) were purchased from Pharmacia LKB Biotechnology, Inc. Other reagents were obtained from Sigma.

Molecular cloning and DNA hybridization techniques

Most of the methods used for molecular cloning were based on those of Sambrook *et al.* (1989). *E. coli* MV1184 was mainly used as a host for plasmid preparations. Chromosomal DNA of *T. filiformis* was isolated by the method of Marmur (1961). Plasmid DNA was prepared by a modified alkaline extraction method (Sambrook *et al.*, 1989). The transformation of *E. coli* was performed as described by Hanahan (1983) and Kushner (1973). DNA was labeled by nick-translation according to Rigby *et al.* (1977). The DNA probe used for the DNA-DNA hybridization to detect the *Tfi* DNA polymerase gene was the 1.8 kb *Hin*dIII fragment from pKTPOL10 containing the *Taq* DNA polymerase gene (Kwon *et al.*, 1991). Agarose gel membrane hybridization was performed by the method of Silhavy *et al.* (1984). Colony hybridization was performed by the method of Hanahan and Meselson (1980).

DNA sequencing and computer-assisted analyses

The restriction fragments to be sequenced were cloned into appropriate restriction sites of pUC18/19 and pBluescript®II SK+/− vectors. DNA sequencing by the dideoxynucleotide chain-termination method was performed according to Hattori and Sakaki (1986) using an alkali-denatured plasmid DNA as the template and universal primer. Sequence data was analyzed using PCGENE and DNASIS as DNA analysis programs.

Results and Discussion

Cloning of the Tfi *DNA polymerase gene*

To clone the *Tfi* DNA polymerase gene, the structural gene coding for *Taq* DNA polymerase was used as a hybridization probe (Kwon *et al.*, 1991). Chromosomal DNA prepared from *T. filiformis* was digested with restriction enzymes, followed by separation by 0.8% agarose gel electrophoresis. The agarose gel was dehydrated, and agarose-membrane hybridization was performed using ^{32}P-labeled *Taq* DNA polymerase gene. The probe hybridized to both an approximately 4.8 kb and a 2.2 kb *Bam*HI fragments, an approximately 12 kb *Hin*dIII fragments, and both an approximately 6 kb and a 4.5 kb *Pst*I fragments (Fig. 1). Two *Bam*HI fragments were suitable for cloning, because the *Bam*HI fragments was smaller than *Pst*I and *Hin*dIII fragments. Accordingly, *T. filiformis* DNA (100 μg) was digested with *Bam*HI and then electrophoresed on a low-melting agarose gel. The resulting DNA fragments were separately collected from the regions containing the 4.8 kb and 2.2 kb *Bam*HI fragments. The 4.8 kb and 2.2 kb *Bam*HI fragments were separately ligated at the *Bam*HI site in the multiple cloning sites of plasmid vector pUC18, and then *E. coli* MV1184 was

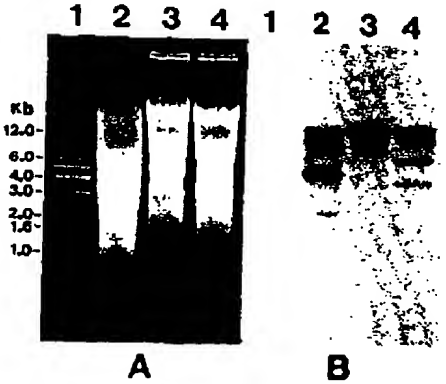

Figure 1. Southern blot analysis of *T. filiformis* DNA digested with restriction endonucleases. The DNA probe used for DNA hybridization is the 1.8 kb *Hin*dIII fragment from pKTPOL10 containing the *Taq* DNA polymerase gene (Kwon *et al.*, 1991). A) EtBr stained agarose gel. B) Southern blot hybridization with DNA probe. Lane 1, DNA molecular weight marker X (0.07-12.2 kbp); lane 2, *Bam*HI-digested genomic DNA; lane 3, *Hin*dIII-digested genomic DNA; lane 4, *Pst*I-digested genomic DNA.

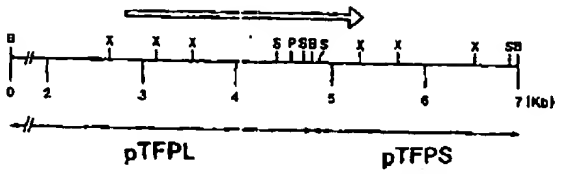

Figure 2. Restriction map and positions of the cloned DNA fragments in plasmid pTFPL and pTFPS. The restriction sites used for the subcloning and DNA sequencing of cloned DNA fragments are shown: B, *Bam*HI; P, *Pst*I; S, *Sac*I; X, *Xho*I. The position of the *Tfi* DNA polymerase gene in the cloned fragments is indicated by the open arrow.

transformed with the plasmids. After colony hybridization, more than 16 clones showed a prominent reaction with the probe. Plasmid DNAs were prepared from these clones, and finally two kinds of plasmids, named pTFPL and pTFPS, were obtained. pTFPL contained the 4.8 kb *Bam*HI fragment to which the labeled probe hybridized, and pTFPS contained the 2.2 kb *Bam*HI fragment to which the labeled probe hybridized (Fig. 2).

Nucleotide sequence of the Tfi *DNA polymerase gene and its deduced amino acid sequence*

The restriction maps of the 4.8 kb and 2.2 kb *Bam*HI fragments are presented in Figure 2. Each enzyme site of the restriction maps was used for the subcloning and DNA sequencing of cloned DNA fragments. Both strands of the subclones containing the gene of *Tfi* DNA polymerase were sequenced. The position of the *Tfi* DNA polymerase gene in the cloned fragments is indicated by the open arrow. Figure 3 shows the nucleotide sequence of the DNA and the deduced amino acid sequence of *Tfi* DNA polymerase. *Tfi* DNA polymerase was comprised of 833 amino acid residues and its molecular mass was determined to be 93,890 Da.

The amino acid composition of *Tfi* DNA polymerase, calculated from the deduced sequence, is shown in Table 1 and is compared with those of other *E. coli* DNA polymerase I-like DNA polymerases. The amino acid composition of *Tfi* DNA polymerase is similar to that of other enzymes. Thermophilic organisms cannot regulate their internal temperature. Consequently, thermophilic organisms must possess

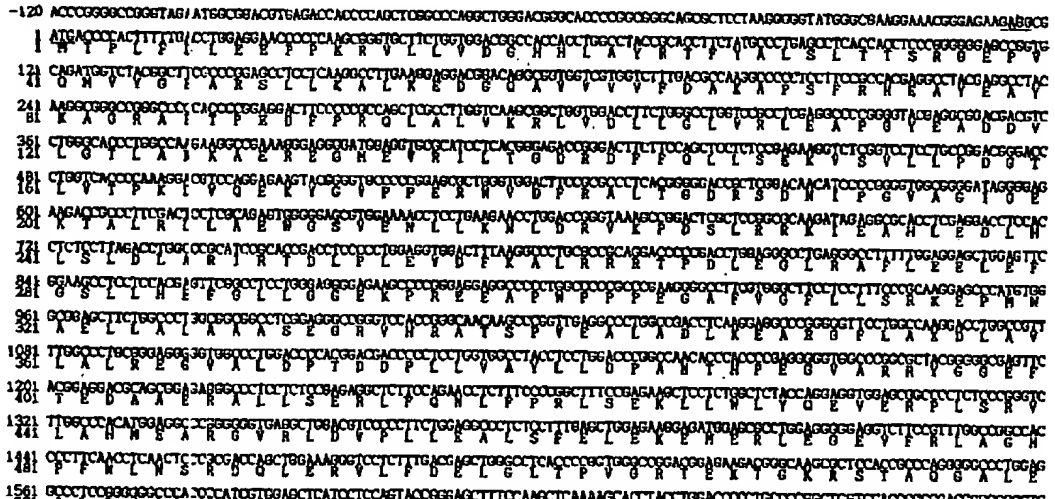

Figure 3. Nucleotide sequence of the cloned DNA fragments and deduced amino acid sequence of *Tfi* DNA polymerase. The numbering of nucleotides starts at the 5'-terminus of the gene encoding *Tfi* DNA polymerase, and that of amino acids start at the NH_2-terminus of *Tfi* DNA polymerase. A putative Shine-Dalgarno sequence is underlined. Asterisks indicate the stop codon.

Table 1. Amino acid composition of *Tfi* DNA polymerase in comparison with those of other DNA polymerases

Amino acid	*Tfi* DNA polymerase	*Taq* DNA polymerase	*Tca* DNA polymerase	*E. coli* DNA polymerase I
Ala	82	91	89	99
Arg	77	76	68	45
Asn	14	12	18	32
Asp	47	42	41	51
Cys	0	0	1	2
Gln	19	16	22	39
Glu	84	87	83	80
Gly	55	58	54	57
His	18	18	21	21
Ile	15	25	16	53
Leu	128	124	127	106
Lys	39	42	45	59
Met	14	16	16	25
Phe	34	27	29	24
Pro	50	48	51	50
Ser	32	31	30	39
Thr	35	30	31	50
Trp	9	14	12	7
Tyr	19	24	20	32
Val	62	51	60	57
Total	833	832	834	928
M_r	93,890	93,922	93,810	103,130

intrinsically thermostable cellular enzymes. A number of reports have been written on the enhanced stability of thermophilic enzymes (Thomas and William, 1986). The thermostability of an enzyme is a basic function of the enzyme's stabilizing forces. These include hydrophobic interactions, disulfide bridges, ionic interactions, hydrogen bonding, and metal binding. In their absence, destabilizing forces arise from the conformational entropy of the protein. Each of these stabilizing forces, either by itself or in combination, has been suggested as a possibility for enhanced thermostability. However, *Tfi* DNA polymerase does not contain a disulfide bridge (Table 1). The ratios of hydrophobic amino acid-composition between *Tfi* DNA polymerase and *E. coli* DNA polymerase I also showed some similarity, but the thermostability of two enzymes was different. Therefore, the thermostability of *Tfi* DNA polymerase cannot be elucidated by comparison of amino acid compositions. *Tfi*, *Taq*, and *Tca* DNA polymerases have lower Lys and higher Arg contents than *E. coli* DNA polymerase I, which is characteristic of enzymes derived from the genus *Thermus* (Kagawa *et al.*, 1984; Kunai *et al.*, 1986).

Comparison of the amino acid sequence of Tfi *DNA polymerase with those of other* E. coli *DNA polymerase I-like DNA polymerases*

The whole amino acid sequence of *Tfi* DNA polymerase showed a high homology to those of the *E. coli* DNA polymerase I-like DNA polymerases *Taq* DNA polymerase (Lawyer *et al.*, 1989), *Tca* DNA polymerase (Kwon *et al.*, 1997), and *E. coli* DNA polymerase I (Joyce *et al.*, 1982) (Fig. 4). Extremely high sequence homology was observed in the *Tfi*, *Taq*, and *Tca* DNA polymerase. *Tfi* DNA polymerase shows 78.5% homology to *Taq* DNA polymerase, 78.4% to *Tca* DNA polymerase, and 41.8% to *E. coli* DNA polymerase I.

In the case of *E. coli* DNA polymerase I, proteolytic cleavage separates the polypeptide chain into two active fragments; a smaller NH_2-terminal fragment containing the 5' → 3' exonuclease activity and a large COOH-terminal fragment that contains polymerase and 3' → 5' exonuclease activities (Derbyshire *et al.*, 1988; Jacobsen *et al.*, 1974; Klenow and Henningsen, 1970; Ollis *et al.*, 1985). The NH_2-terminal regions of *Tfi*, *Taq*, and *Tca* DNA polymerase correspond to the NH_2-terminal domain of *E. coli* DNA polymerase I. In *Tfi* DNA polymerase, the first 254 amino acids from the NH_2 terminus showed homology to the 5' → 3' exonuclease domain of *E. coli* DNA polymerase I. In agreement with this structural data, *Tfi*, *Taq*, and *Tca* DNA polymerases exhibit 5' → 3' exonuclease activity. The COOH-terminal regions of *Tfi*, *Taq*, and *Tca* DNA polymerase correspond to that of the *E. coli* DNA polymerase I containing DNA polymerase activity. As shown in Figure 4, this region is conserved in most of the DNA polymerases, suggesting that this region corresponds to an evolutionarily conserved DNA polymerase domain (Blanco *et al.*, 1991).

As a result of mutations, deletions, and substitutions during evolution, *Tfi* DNA polymerase residues at positions 255-433 show little sequence similarity to the *E. coli* DNA polymerase I domain (at positions 261-529) assumed to contain the 3' → 5' exonuclease activity (Derbyshire *et al.*, 1988; Ollis *et al.*, 1985). In this region, the amino acid sequence of *Tfi* DNA polymerase showed a especially high homology to those of *Taq* and *Tca* DNA polymerase, but *E. coli* DNA polymerase I had a highly different structure and showed little similarity to the others. *Tfi* DNA polymerase is 95 residues shorter than *E. coli* DNA polymerase I because most of the deleted residues occur in the region encompassing residues 255-433. In *E. coli* DNA polymerase, this domain structurally contains the 3' → 5' exonuclease active site. A common feature of many DNA polymerases is a 3' → 5' exonuclease activity that is partly responsible for the high fidelity of DNA replication (Kunkel, 1988). This evolutionarily conserved active site is mainly formed by the highly conserved regions *Exo*I, *Exo*II, and *Exo*III (Blanco *et al.*, 1991) as shown in Figure 4. However, the regions of *Tfi* DNA polymerase corresponding to the highly conserved regions *Exo*I, *Exo*II, and *Exo*III of *E. coli* DNA polymerase I did not exist. Therefore, it is reasonable to believe that *Tfi*, *Taq*, and *Tca* DNA polymerases do not possess as much 3' → 5' exonuclease activity, as *E. coli* DNA

polymerase I. Actually, 3' → 5' exonuclease activity cannot be detected in the purified *Tca* DNA polymerase (Park *et al.*, 1993).

Analysis of 5'- and 3'-noncoding regions of the Tca DNA polymerase gene

Analysis of the gene that codes for *Tfi* DNA poly-

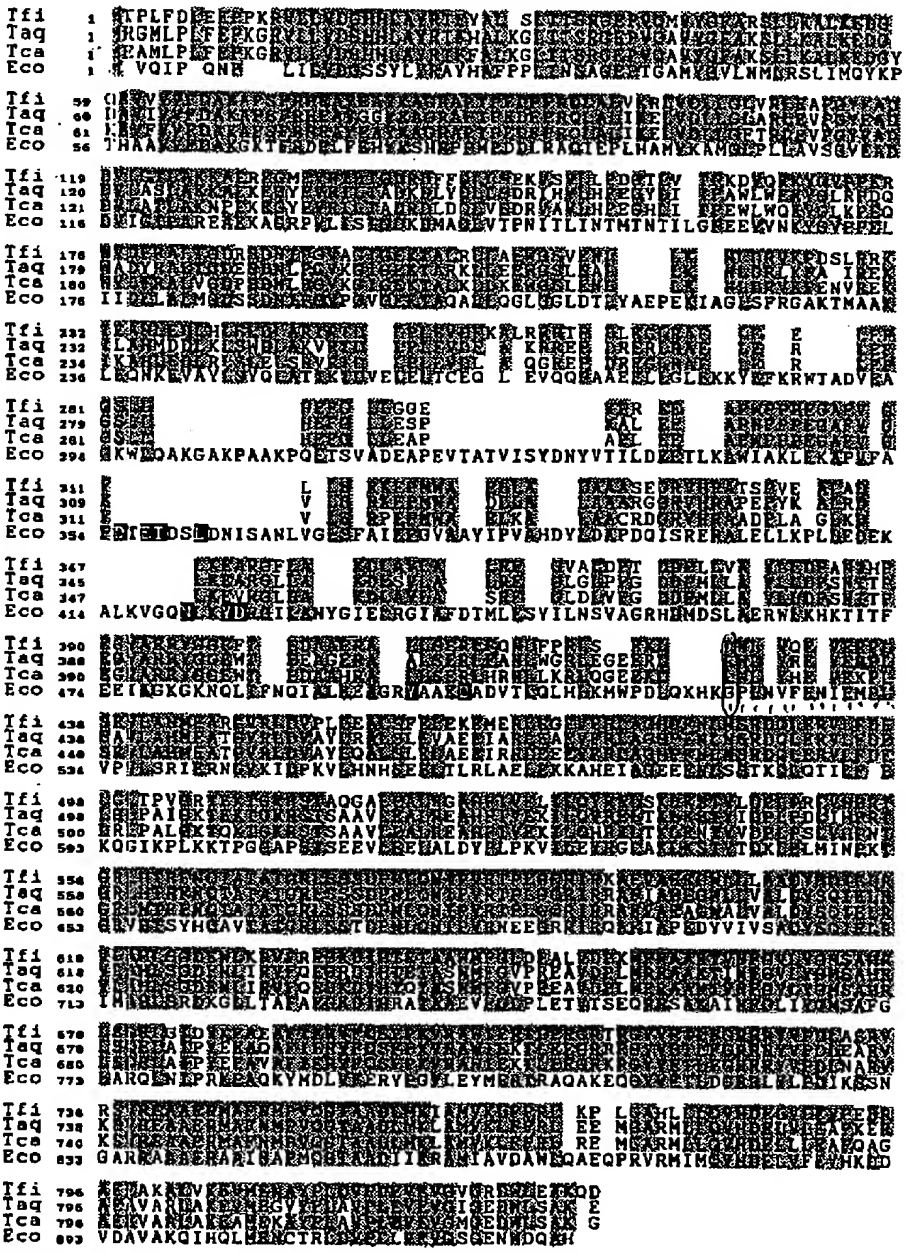

Figure 4. Comparison of the amino acid sequence of *Tfi* DNA polymerase with those of other *E. coli* DNA polymerase I-like DNA polymerases. The sequence of *Tfi* DNA polymerase (Tfi) is shown as compared with those of *Taq* DNA polymerase (Taq) (Lawyer *et al.*, 1989), *Tca* DNA polymerase (Tca) (Kwon *et al.*, 1997), and *E. coli* DNA polymerase (Eco) (Joyce *et al.*, 1982). Identical amino acids between *Tfi* DNA polymerase and others are indicated by stippled boxes. Dark-shaded boxes indicate the three highly conserved regions *Exo*I, *Exo*II, and *Exo*III, proposed to form a general 3' → 5' exonuclease-active site (Blanco *et al.*, 1991).

merase reveals other interesting features. In the 5'-noncoding region of the gene, the 5'-GAGG-3' segment at position −6 to −3 upstream of the translation start codon (ATG for Met) complements the 3' end of the 16S rRNA of *E. coli* (Shine and Dalgarno, 1975) and resembles a ribosome binding site.

Except for the promoter sequence of the *T. flavus* succinyl-coenzyme A synthetase-malate dehydrogenase (Nishiyama *et al.*, 1991), the promoter-like sequence in the −35 region and the −10 region, which can function in *E. coli*, was not found in the most of the genes derived from genus *Thermus* (Kunai *et al.*, 1986; Kwon *et al.*, 1997; Lawyer *et al.*, 1989). The promoter-like sequence was not also found in the upstream flanking region of the *Tfi* DNA polymerase gene.

In the 3'-noncoding region of the gene, there was no potential transcriptional termination sequence able to form a stem-and-loop structure followed by a pyrimidine-rich sequence (Fig. 3).

High G+C content of the third positions in the codon usage

The G+C content of the *Tfi* DNA polymerase gene was 68.5%, slightly higher than that of chromosomal DNA (65%) (Hudson *et al.*, 1987). The G+C content in the first, second, and third positions of the codons used were 78.1%, 40.9%, and 92.7% respectively. Codon usage in *Tfi* DNA polymerase was heavily biased towards the use of G and C in the third position, as expected for an organism with G+C rich DNA (Table 2). Essentially identical third-position codon bias has been observed for other *Thermus* genes: 93% G+C in third position for the *Tca* DNA polymerase gene of *T. caldophilus* GK24 (Kwon *et al.*, 1997), 91.8% for the *Taq* DNA polymerase gene of *T. aquaticus* YT-1 (Lawyer *et al.*, 1989), 94.8% for malate dehydrogenase gene of *T. flavus* AT-62 (Nishiyama *et al.*, 1986), and 89.4% for the isopropylmalate dehydrogenase gene of *T. thermophilus* HB8 (Kagawa *et al.*, 1984). Codons, the third positions of which are U or A, are thus rarely used, and only 61 such codons were observed among 834 codons in the *Tfi* DNA polymerase gene.

The codons of Arg, Ala, Pro, and Gly generally raise the G+C content of DNA, while those of Lys, Ile, Met, Tyr, Asn, and Phe raise the A+T content (Table 2). The amino acid compositions of DNA polymerases from *T. filiformis* and *E. coli* are shown in Table 1. Specially, Arg content was much higher in the *Tfi* DNA polymerase gene than in the *E. coli* DNA polymerase I gene. On the other hand, Lys, Ile, Met, Tyr, and Asn levels were much lower in the *Tfi* DNA polymerase gene than in the *E. coli* DNA polymerase I gene. These changes in amino acid composition increase the G+C content of the DNA.

Special codon usage avoiding the GA(A/T)TC (TfiI site) and TCGA sequence (TaqI site)

Chromosomal DNA from *T. filiformis* was not digested by *Tfi*I restriction endonuclease at all. There was no nucleotide sequence of GA(A/T)TC (*Tfi*I recognition site) in the sequenced region of *T. filiformis* DNA (2,502 nucleotides) (Fig. 3). There was also no nucleotide sequence of TCGA (*Taq*I recognition site) except for three nucleotide sequences of CTCGAG

Table 2. Codon usage in the gene for *Tfi* DNA polymerase in comparison with that in the gene for *Taq* and *Tca* DNA polymerases

1st base	2nd base																3rd base
	U				C				A				G				
		Tfi	Taq	Tca		Tfi	Taq	Tca		Tfi	Taq	Tca		Tfi	Taq	Tca	
U	Phe	8	8	5	Ser	0	0	0	Tyr	2	4	2	Cys	0	0	0	U
	Phe	26	19	24	Ser	18	15	16	Tyr	17	20	18	Cys	0	0	1	C
	Leu	2	0	0	Ser	0	0	0	Trm	0	0	0	Trm	1	1	0	A
	Leu	8	3	8	Ser	6	1	2	Trm	0	0	1	Trp	9	14	12	G
C	Leu	11	20	14	Pro	0	3	0	His	1	0	2	Arg	1	0	0	U
	Leu	54	46	72	Pro	36	34	39	His	17	18	19	Arg	26	24	22	C
	Leu	1	5	3	Pro	3	2	2	Gln	1	1	4	Arg	1	0	0	A
	Leu	52	50	30	Pro	11	9	10	Gln	18	15	18	Arg	36	27	25	G
A	Ile	2	3	0	Thr	0	0	0	Asn	0	0	1	Ser	0	1	1	U
	Ile	9	20	13	Thr	22	20	22	Asn	14	12	17	Ser	8	14	11	C
	Ile	4	2	3	Thr	1	0	0	Lys	1	5	2	Arg	0	0	1	A
	Met	14	16	16	Thr	12	10	9	Lys	38	37	43	Arg	13	15	16	G
G	Val	2	0	1	Ala	1	0	2	Asp	2	1	3	Gly	1	0	2	U
	Val	25	25	21	Ala	62	68	77	Asp	45	40	39	Gly	23	28	23	C
	Val	1	2	0	Ala	3	2	0	Glu	6	10	8	Gly	5	0	2	A
	Val	34	33	29	Ala	16	19	12	Glu	78	73	79	Gly	26	30	27	G

Tfi, *Tfi* DNA plymerase; Taq, *Taq* DNA polymerase; Tca, *Tca* DNA polymerase.

(*Xho*I site) and one nucleotide sequence of TTCGAC in the sequenced region of *T. filiformis* DNA (2,502 nucleotides) (Fig. 3) TCGA is the recognition sequence for the restriction endonucleases *Taq*I and *Tth*HB8I (Barany *et al.*, 1992), which have been purified from *T. aquaticus* YT-1 and *T. thermophilus* HB8, respectively. The *Taq*I site was not present in the DNA sequences of various *Thermus* chromosomes (Kunai *et al.*, 1986; Kwon *et al.*, 1988). Chromosomal DNA from *T. filiformis* was not also digested by *Taq*I and *Tth*HB8I at all, suggesting that *T. filiformis* has the same host restriction and modification system as other *Thermus* species. A DNA adenine methylase from *T. thermophilus* HB8 has been reported (Sato *et al.*, 1980). This enzyme recognizes sequences of CTCGAG and TTCGAC in *Thermus* cells, and the methylated sequence of TCG^{m}A cannot be used as a substrate for *Taq*I.

We have examined the numbers of NCGA, TNGA, TCNA, and TCGN sequences, which are sequences similar to TCGA (*Taq*I recognition site). In the sequenced region of *T. filiformis* DNA (Fig. 3), the numbers were 33, 53, 35, and 26 respectively. We have also examined the numbers of NA(A/T)TC, GN(A/T)TC, GANTC, GA(A/T)NC, and GA(A/T)TN sequences, which are sequences similar to GA(A/T)TC (*Tfi*I recognition site) In the sequenced region of the *T. filiformis* DNA (Fig. 3), the numbers were 0, 11, 4, 11, and 2, respectively. This data suggests that the TCGA and GA(A/T)TC sequences are avoided in *T. filiformis*.

Avoiding the sequences of GA(A/T)TC and TCGA sometimes results in the usage of rare codons (A or T in the third position) in the *Tfi* DNA polymerase gene. There were only 22 rare codons, the third bases of which are U or A, in the *Tfi* DNA polymerase gene (Fig. 3, Table 2). We are conducting experiments to express the *Tfi* DNA polymerase gene in *E. coli*.

Acknowledgment

This study was supported by the Academic Research Fund (GE96-216) of the Ministry of Education, Republic of Korea.

References

Alting-Mees, M. A., and Short, J. M. (1989) *Nucleic Acids Res.* **17**, 9494.

Barany, F., Danzitz, M., Zebala, J., and Mayer, M. (1992) *Gene* **112**, 3-12.

Bernad, A., Blanco, L., Lazaro, J. M., Martin, G., and Salas, M. (1989) *Cell* **59**, 219-228.

Blanco, L., Bernad, M. A., and Salas, M. (1991) *Gene* **100**, 27-38.

Chien, A., Edgar, D. B., and Trela, J. M. (1976) *J. Bacteriol.* **127**, 1550-1557.

Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M., and Steitz, T. A. (1988) *Science* **240**, 199-201.

Erlich, H. A. (1989) *PCR Technology: Principles and Applications for DNA Amplification*, M Stockton Press, New York.

Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557-580.

Hanahan, D., and Meselson, M. (1980) *Gene* **10**, 63-67

Hattori, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232-238.

Hudson, J. A., Morgan H. W., and Daniel, R. M. (1987) *Int. J. Syst. Bacteriol.* **37**, 431-436.

Ito, J., and Braithwaite, D. K. (1991) *Nucleic Acids Res.* **19**, 4045-4057.

Jacobsen, H., Klenow, H., and Vergaard-Hansen, K. (1974) *Eur. J. Biochem.* **45**, 623-627.

Joyce, C. M., Kelly, W. S., and Grindley, N. D. F. (1982) *J. Biol. Chem.* **257**, 1958-1964.

Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T., and Oshima, T. (1984) *J. Biol. Chem.* **259**, 2956-2960.

Kaledin, A. S., Slyusarenko, A. G., and Gorodetskii, S. Z. (1980) *Biokhimiya* **46**, 1576-1584.

Kaledin, A. S., Slyusarenko, A. G., and Gorodetskii, S. Z. (1982) *Biokhimiya* **47**, 1758-1791.

Klenow, H., and Henningsen, I. (1970) *Proc. Natl. Acad. Sci. USA* **65**, 168-175.

Kunai, K., Machida, M., Matsuzawa, H., and Ohta, T. (1986) *Eur. J. Biochem.* **160**, 433-440.

Kunkel, T. A. (1988) *Cell* **53**, 837-840.

Kushner, S. R. (1978) in *Genetic Engineering* (Boyer, H. B., and Nicosia, S., eds) pp. 17-23, Elsevier/North Holland, Amsterdam.

Kwon, S.-T., Terada, I., Matsuzawa, H., and Ohta, T. (1988) *Eur. J. Biochem.* **173**, 491-497.

Kwon, S.-T., Kim, J. S., Park, J. H., Koh, S. H., and Lee, D.-S. (1991) *Mol. Cells* **1**, 369-375.

Kwon, S.-T., Kim, J. S., Park, J. H., Kim, H. K., and Lee, D.-S. (1997) *Mol. Cells* **7**, 264-271.

Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. (1989) *J. Biol. Chem.* **264**, 6427-6437.

Lopez, P., Martinez, S., Diaz, A., Espinosa, M., and Lacks, S. A. (1989) *J. Biol. Chem.* **264**, 4255-4263.

Marmur, J. (1961) *J. Mol. Biol.* **3**, 208-218.

Nishiyama, M., Matsubara, N., Yamamotor, K., Iijima, S., Uozumi, T., and Beppu, T. (1986) *J. Biol. Chem.* **261**, 14178-14183.

Norrander, J., Kempe, T., and Messing, J. (1983) *Gene* **26**, 101-106.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) *Nature* **313**, 762-766.

Park, J. H., Kim, J. S., Kwon, S.-T., and Lee, D.-S. (1993) *Eur. J. Biochem.* **214**, 135-140.

Ramaley, R. F., and Hixson, J. (1970) *J. Bacteriol.* **103**, 527-528.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113**, 237-251.

Ruettimann, C., Cotoras, M., Zaldivar, J., and Vicuna, R. (1985) *Eur. J. Biochem.* **149**, 41-46.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., and Erlich, H. A., and Arnheim, N. (1985) *Science* **230**, 1350-1354.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. Higuchi, R., Horn, G. T., and Erlich, H. A. (1988) *Science*

239, 487-491.
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed, pp. A9-A13, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sato, S., Nakazawa, K., and Shinomiya, T. (1980) *J. Biochem.* **88**, 737-747.
Shine, J., and Dalgarno, L. (1975) *Nature* **254**, 34-38.
Short, J. M., Fernandez, J. M., Sorge, J. A., and Huse, W. D. (1988) *Nucleic Acids Res.* **16**, 7583-7600.
Silhavy, T. J., Berman, M. L., and Enquist, L. W. (1984) *Experiments with Gene Fusions*, pp. 189-195, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Thomas, K. N., and William, R. K. (1986) in *Thermophiles* (Thomas, D. B., ed) pp. 197-215, John Wiley & Sons, New York.

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☑ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.